

Distributed Kalman filtering compared to Fourier domain preconditioned conjugate gradient for laser guide star tomography on extremely large telescopes

Luc Gilles,^{1,*} Paolo Massioni,² Caroline Kulcsár,³ Henri-François Raynaud,³ and Brent Ellerbroek¹

¹Thirty Meter Telescope Observatory Corp., 1200 E. California Blvd., MC 102-8, Pasadena, California 91125, USA

²Laboratoire Ampère, UMR CNRS 5005, INSA de Lyon, 69621 Villeurbanne CEDEX, France

³Laboratoire Charles Fabry de l'Institut d'Optique—CNRS—Université Paris-Sud Orsay, RD 128, 91127 Palaiseau CEDEX, France

*Corresponding author: lgilles@caltech.edu

Received January 31, 2013; revised March 6, 2013; accepted March 6, 2013;
posted March 7, 2013 (Doc. ID 178608); published April 16, 2013

This paper discusses the performance and cost of two computationally efficient Fourier-based tomographic wavefront reconstruction algorithms for wide-field laser guide star (LGS) adaptive optics (AO). The first algorithm is the iterative Fourier domain preconditioned conjugate gradient (FDPCCG) algorithm developed by Yang *et al.* [Appl. Opt. **45**, 5281 (2006)], combined with pseudo-open-loop control (POLC). FDPCCG's computational cost is proportional to $N \log(N)$, where N denotes the dimensionality of the tomography problem. The second algorithm is the distributed Kalman filter (DKF) developed by Massioni *et al.* [J. Opt. Soc. Am. A **28**, 2298 (2011)], which is a noniterative spatially invariant controller. When implemented in the Fourier domain, DKF's cost is also proportional to $N \log(N)$. Both algorithms are capable of estimating spatial frequency components of the residual phase beyond the wavefront sensor (WFS) cutoff frequency thanks to regularization, thereby reducing WFS spatial aliasing at the expense of more computations. We present performance and cost analyses for the LGS multiconjugate AO system under design for the Thirty Meter Telescope, as well as DKF's sensitivity to uncertainties in wind profile prior information. We found that, provided the wind profile is known to better than 10% wind speed accuracy and 20 deg wind direction accuracy, DKF, despite its spatial invariance assumptions, delivers a significantly reduced wavefront error compared to the static FDPCCG minimum variance estimator combined with POLC. Due to its nonsequential nature and high degree of parallelism, DKF is particularly well suited for real-time implementation on inexpensive off-the-shelf graphics processing units. © 2013 Optical Society of America

OCIS codes: (010.1080) Active or adaptive optics; (010.1330) Atmospheric turbulence.

<http://dx.doi.org/10.1364/JOSAA.30.000898>

1. INTRODUCTION

Tomographic wavefront reconstruction using laser guide stars (LGSs) is under development for adaptive optics (AO) systems on extremely large telescopes (ELTs) [1–4]. Due to LGS angular and range uncertainty [5,6], a few dim natural guide stars (NGSs) are still required to supplement the high-order LGS wavefront sensor (WFS) measurements, to estimate a few low-order tomographic modes poorly measured by the LGS WFSs. The overall wavefront reconstruction problem is thus traditionally split into separate high- and low-order reconstruction problems, in an architecture known as “split tomography” [7–9]. The low-order problem is low-dimensional and does not present any computational bottleneck, whereas the high-order problem is computationally challenging, since its dimensionality grows as the square of the telescope diameter. Following tomographic wavefront reconstruction, a deterministic least-squares fitting of this estimate is performed onto deformable mirror (DM) actuators [10]. When DM actuators are distributed on a single DM, the configuration is known as laser tomography AO (LTAO), whereas when the projection is performed onto multiple DMs in series at different conjugate ranges, the configuration is known as

multiconjugate AO (MCAO). Both configurations operate with closed-loop feedback. If the DM(s) are controlled open-loop (i.e., their shapes are invisible to the WFSs), the configuration is known as multiobject AO (MOAO). For a system like the narrow-field infrared AO (NFIRAOS) system, which is the LGS MCAO system under design for the Thirty Meter Telescope (TMT) project [3], a total of $N_g \sim 34,000$ LGS WFS measurements from $N_{\text{lgs}} = 6$ LGS WFSs are used to reconstruct $N_{\text{ps}} = 6$ atmospheric layers, which are subsequently projected onto $N_a \sim 7,000$ DM actuators distributed on $N_{\text{dm}} = 2$ DMs.

A large amount of work has been performed in the last few years to develop computationally efficient algorithms and techniques to solve the challenging tomography step. The algorithms investigated fall into two categories: iterative and noniterative. Iterative solutions include conjugate gradient (CG) [8], Fourier domain preconditioned CG (FDPCCG) [11,12], a fractal iterative method [13,14], and a Kaczmarz-type reconstructor [15]. Noniterative solutions include structured Kalman filtering [16], distributed Kalman filtering (DKF) [17], and massively parallel matrix vector multiplication (MVM), directly mapping measurements to actuators [18]. With the advent of more

powerful, off-the-shelf, cost-effective graphics processing units (GPUs), it becomes conceivable to implement in real time all of those solutions for ELT-scale problems on such computing engines. A specialized AO tomography workshop was held recently to brainstorm on those techniques [19].

This paper analyzes the performance and computational cost of the FDPCG and DKF algorithms. The analysis is presented in the context of the LGS MCAO system under design for TMT. Key differences between both algorithms are the following.

- DKF is a spatially invariant controller; i.e., it assumes an infinite telescope pupil and uniform measurement noise, which leads to a controller that consists of $2N_{\text{ps}}N_{\text{LGS}}$ spatial domain (SD) convolution operators, each blur kernel of size $n_K \times n_K$. On the other hand, FDPCG takes into account the finite telescope pupil, partially illuminated subapertures, and nonuniform measurement noise arising from LGS spot elongation. For the TMT simulations discussed in this paper, we found that $n_K = D/d_{\text{sa}} = 60$ was required in order to reach best performance, where $D = 30$ m denotes the telescope diameter and $d_{\text{sa}} = 1/2$ m is the LGS WFS subaperture size. When phase points are estimated at twice the subaperture resolution (we will refer to such layers as “oversampled”), we found that $n_K = 2D/d_{\text{sa}} = 120$ was required. Oversampling has the benefit of reducing WFS spatial aliasing at the expense of more computations.

- Regarding computational cost, FDPCG complexity scales as $N \log(N)$, where N denotes the dimensionality of the tomography problem, whereas the SD implementation of DKF as proposed in [17] scales as $N_{\text{ps}}n_K^2N_g$. Since N_g is proportional to N and $N_{\text{ps}}n_K^2$ to N as well (given the required blur size discussed in the previous bullet), this leads to a cost scaling as N^2 . Alternatively, the convolutions can be performed in the Fourier domain (FD), which leads to a more efficient cost scaling as $N \log(N)$. A detailed cost breakdown is discussed as part of our analysis.

- FDPCG is intrinsically a static controller, but it can be extended to include a prediction step [20,21], whereas DKF intrinsically incorporates turbulence dynamics. We will show that the choice of a turbulence temporal model plays a critical role in the ultimate performance delivered by the DKF algorithm. We have chosen a first-order auto-regressive (AR1) frozen flow model (i.e., the evolution operator is a shift operator) as a prediction step for both algorithms, and sensitivity to wind profile prior information is investigated.

It should be pointed out that both algorithms need to be updated at a low rate as part of a background process (typically every few seconds) as turbulence/observing conditions change and the telescope pupil obscuration rotate. Finally, we should note that besides computational cost, memory access, inter-processor communication, and parallelization are additional key factors that need to be considered to determine the most cost-effective hardware architecture to implement these algorithms. Due to its overall simplicity and massively parallel nature, DKF's SD and FD implementations appear in this regard to be well suited for real-time implementation on inexpensive off-the-shelf GPUs.

We found that, provided the wind profile is known to better than 10% wind speed accuracy and 20 deg wind direction accuracy, the DKF algorithm, despite its spatial invariance

assumptions, delivers a significantly reduced wavefront error (WFE) compared to the static FDPCG minimum variance estimator (MVE) with pseudo-open-loop control (POLC). Moreover, this potential performance gain comes at reduced computational cost, as we find that the number of operations for three iterations of FDPCG (which is the required number of iterations to reach convergence and hence best performance) is significantly higher than that of the DKF algorithm with the Kalman gain multiplication implemented via fast Fourier transforms (FFTs). At zenith and under median Mauna Kea seeing (0.55 arcsec at 500 nm wavelength), the performance gain assuming perfect knowledge of the wind profile is estimated to be around 38 nm RMS in quadrature for the TMT LGS MCAO system, which corresponds to an almost +4% Strehl ratio increase in the J band (1.25 μm wavelength), and it is still around 7 nm for a combined 20 deg wind direction error and 10% wind speed error on all estimated layers.

The paper is organized as follows. Section 2 forms the core of the discussion and provides a detailed description and computational cost and performance analysis of the FDPCG and DKF algorithms. A graphic illustration of an efficient implementation of the FD preconditioning step is also discussed, which is a critical component of the FDPCG algorithm. Conclusions are drawn in Section 3.

2. FDPCG AND DKF ALGORITHM DESCRIPTION, COMPUTATIONAL COST, AND PERFORMANCE ANALYSIS

Two split atmospheric tomography control architectures using LGSs and NGSs have been discussed by Gilles *et al.* [7,8]. The main benefits of these split control approaches is (i) a simpler formulation of the LGS high-order tomography step with reduced computations, (ii) a simpler and more flexible control of the low-order NGS-controlled tip/tilt (TT) and tilt anisoplanatism modes, and (iii) reduced coupling between the LGS- and NGS-controlled modes. This paper is concerned with the FDPCG and DKF algorithms to solve the computationally challenging high-order LGS tomography problem. The tomographic estimation problem involves operations on three types of phase grids: (i) N_{ps} atmospheric phase screens conjugated to different altitudes in the atmosphere, (ii) an aperture-plane grid, and (iii) N_{dm} DM actuator grids conjugated to the DMs. In order to facilitate grid-based computations on operator stencils (which reduce storage requirements and memory access compared to sparse matrix implementations), it is convenient to mesh these grids over square SDs that enclose the union of the multiple LGS and science metapupils. In order to keep the number of operations minimal, these square grids can be chosen smaller than the minimal FFT grid size (twice the telescope diameter, $2D$). Grid extension (zero padding) is then required before the forward FFTs, and grid restriction is required after the backward FFTs.

All grids employ Cartesian coordinates, and sampling is uniform across a given grid. The aperture-plane grid samples the LGS WFS subapertures at corner, midvertex, and center locations, which allows us to discretize the WFS measurement operator via Simpson's rule. We will denote by $\Delta_0 = d_{\text{sa}}/2$ the aperture-plane mesh size. In order to accommodate the finite-range LGS cone effect in the FD, it is necessary to define the atmospheric screens in cone coordinates [11], with sampling given by $\Delta_k = \alpha_k \xi_k \Delta_0$, where $\xi_k = 1 - h_k/H \leq 1$ denotes the

cone compression factor for a layer at altitude h_k and an LGS at altitude H , and $\alpha_k = 2$ or 1 depending on whether phase points are estimated at the WFS subaperture resolution or at twice that resolution (which is possible thanks to the regularization term appearing in the estimator that incorporates turbulence statistics prior information). We will refer to atmospheric grids with $\alpha_k = 1$ as “oversampled.” With these definitions, the number of oversampled atmospheric layers is given by $N_{os} = 2N_{ps} - \sum_{k=1}^{N_{ps}} \alpha_k$. Oversampling is typically implemented on the strongest turbulence layers, although from an implementation standpoint, it may be easier to oversample all layers. Finally, we will denote by $n_f \geq 2D/d_{sa}$ the 1D number of FFT grid points on nonoversampled layers, and by $2n_f$ that on oversampled ones. Note that n_f is also equal to the 1D number of FFT grid points covering the $1/d_{sa}$ WFS bandwidth.

We assume the following linear “plant” (closed-loop WFS measurement) model:

$$s_n = \Gamma \tilde{H} x_{n-1} - G_a a_{n-1} + \eta_{n-1}, \quad (1)$$

where s_n denotes the closed measurement vector available after detector readout somewhere in the time interval $T[n-1, n]$ (with T denoting the sampling period), $x_{n-1} = (1/T) \int_{n-2}^{n-1} x(t) dt$ denotes the unknown zero-mean concatenated turbulence vector during the WFS integration time interval $T[n-2, n-1]$, Γ is the block-diagonal sparse aperture-plane WFS gradient matrix, \tilde{H} is the block-structured sparse LGS-to-aperture-plane ray-tracing matrix (bilinear interpolations), G_a is the block-structure sparse DM interaction matrix, a_{n-1} is the DM command applied during WFS integration, and η_{n-1} is the additive photon and readout noise vector. Block Γ_i of Γ (with concatenated components $[\Gamma_{xi}]_i, [\Gamma_{yi}]_i$) is characterized by a 3×3 stencil; i.e., row j of $[\Gamma_{xy}]_i$ has up to nine nonzero values corresponding to the eight grid points bordering subaperture j plus the central point. For a perfectly aligned system, all blocks Γ_i of Γ are identical. Block \tilde{H}_{ik} provides phase values on layer k at intercepts of rays traced from LGS i to the aperture plane. Interaction matrix G_a is composed of blocks $[G_{a,il}]$ for WFS i and DM l , each with concatenated components $[G_{a,x}]_{il}, [G_{a,y}]_{il}$. Both FDPCG and DKF make use of the following “pseudo-open-loop” quantity:

$$s_n^{OL} = s_n + G_a a_{n-1}. \quad (2)$$

A. FDPCG

Iterative CG-based solvers to tackle the static open-loop MVE problem arising in atmospheric tomography have been discussed in the past (see, e.g., [7] and references therein). The block-structured system to be solved is expressed as follows:

$$Ax_{n-1} = b_n, \quad (3)$$

where A denotes the forward (or coefficient) block-structured (symmetric) tomography matrix, and b_n the right-hand-side phase vector computed from Eq. (2). Prior information appearing in the definition of A and b_n includes (i) the block-diagonal spatial covariance matrix C_x of the turbulence (assumed to follow von Karman statistics) and (ii) the inverse spatial covariance matrix W of the measurement noise, obtained by postprocessing the subaperture time average pixel intensities and centroider gain (see, e.g., [22]

for a matched filter centroiding algorithm). We consider the FDPCG algorithm [11,12] running in the SD with the preconditioned step implemented via FFTs as the iterative solver for Eq. 3. In such an implementation, following Ellerbroek [23], the block-diagonal turbulence inverse covariance matrix is approximated by a scaled Laplacian squared matrix with periodic boundary conditions, $C_x^{-1} \sim L^2$, where each block of L is characterized by the standard five-point 3×3 Laplacian stencil. The Laplacian scaling is chosen such that $\langle x_n^T C_x^{-1} x_n \rangle = \langle x_n^T L^2 x_n \rangle$, where $\langle \cdot \rangle$ denotes ensemble average. Its derivation is provided in Appendix A.

The algorithm is initialized following the steps described in Table 1. Line 5 corresponds to global TT removal from the measurements on account of the poor TT measurement by LGS WFSs. n_{sa} denotes the number of active subapertures per WFS. The right-hand-side vector b_n (line 6) is assembled by applying the noise weighting matrix (inverse of the measurement noise covariance matrix) to the TT removed pseudo-open-loop measurement, mapping to WFS pupil-plane phase space by multiplying by the transpose of the WFS gradient matrix, and backpropagating to the N_{ps} atmospheric grids via the transpose of the bilinear interpolation matrix.

The algorithm is initialized with the previous frame solution (warm restart, line 7), and the initial tomography residual is computed as described in line 8, which involves a key A -mult operation (block-structured coefficient matrix). Note that all atmospheric layers contribute to the volume turbulence mapping to any given layer through forward and backward propagations. The initial search direction is set equal to the initial residual (line 19). The other key initialization step is the application of the FD preconditioner. Mathematically, the preconditioning step corresponds to a B -mult operation, $z = Br$, where $B = A'^{-1}$ is the inverse of a symmetric positive definite approximation A' of A . The FD preconditioner implements this step via $2N_{ps}$ FFTs:

$$[z]_k = \mathcal{F}_k^{-1}[\hat{r}]_k, \quad \hat{r} := \hat{B} \hat{r}, \quad (4)$$

where \mathcal{F}_k denotes the 2D FFT matrix for layer k , and $[\hat{B}]_{kl} = \mathcal{F}_k[B]_{kl} \mathcal{F}_l^{-1}$ denotes the Fourier representation of block k, l of the preconditioning matrix B . The cross coupling of atmospheric layers through ray tracing, and the cross coupling of four spatial frequency points on each oversampled layer (three points outside the WFS spatial bandwidth alias into a point within the bandwidth) suggest that each row of \hat{B} contains at most n_b nonzero values, where n_b is entirely determined by the number of reconstructed layers and by the number of oversampled layers, and is given by

$$n_b = 4N_{os} + (N_{ps} - N_{os}). \quad (5)$$

Moreover, the cross-coupling analysis indicates that \hat{B} can be reordered into a block-diagonal matrix, with small blocks of size $n_b \times n_b$ each:

$$\hat{B} = \sum_{j=1}^{n_f^2} P_j^T \hat{M}_j P_j, \quad (6)$$

where permutation block P_j has n_b rows, each with n_b nonzero elements. Since B is symmetric, the blocks \hat{M}_j are Hermitian; i.e., their diagonal is real valued and their lower triangular part is the complex conjugate of their upper triangular part. Substituting Eq. (6) into (4) leads to

Table 1. FDPCG Tomography Algorithm Initialization Steps

4	FDPCG Computations	Goal is to compute turb. est \hat{x}_{n-1} and subsequently a_{n+1} from s_n^{OL}
5	$[\tilde{s}_{x(y),n}]_i = [s_{x(y),n}^{\text{OL}}]_i - [\omega_{x(y)}]_i^T [s^{\text{OL}}]_i$	global tip/tilt removal with gradient weights $[\omega_{x(y)}]_i$
6	$[b_n]_k = \sum_{i=1}^{N_{\text{ps}}} \tilde{H}_{ik}^T \Gamma_i^T W_i [s_n^{\text{OL}}]_i$	tomography right-hand-side vector
7	$x = \hat{x}_{n-2}$	initial guess is previous frame estimate
8	$[r]_k = [b_n]_k - [Ax]_k = [b_n]_k - \sum_{l=1}^{N_{\text{ps}}} [A]_{kl} x_l$ $[A]_{kl} = \sum_{i=1}^{N_{\text{ps}}} [\tilde{H}_{ik}^T \Gamma_i^T W_i \Gamma_i \tilde{H}_{il} + \delta_{k,l} L_k^2]$	initial residual (A -mult) interpolation offsets (β_x, β_y) in unit of mesh size and measured from lower left nearest neighboring grid pt
	stencil(L_k) = $c_k \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, $c_k = 0.24 \frac{2\pi}{\lambda} (\frac{r_{0,k}}{\Delta_k})^{5/6}$	Γ_i has irregular 3×3 stencil (18 weights per subap)
	stencil($[\tilde{H}]_{il}$) = $\begin{bmatrix} (1-\beta_x)\beta_y & \beta_x\beta_y \\ (1-\beta_x)(1-\beta_y) & \beta_x(1-\beta_y) \end{bmatrix}$	Laplacian has regular 3×3 stencil (five nonzero weights)
13	$z = \text{Precon}(r, [\hat{M}]_1, \dots, [\hat{M}]_{n_f^2})$	preconditioning step
14	$[\hat{r}]_k = \mathcal{F}_k[r]_k$ ($k = 1, \dots, N_{\text{ps}}$)	2D forward FFTs for each layer
15	read n_b values $[\hat{r}]_b = P_j \hat{r}$ ($j = 1, \dots, n_f^2$) from \hat{r}	read 4 pts on each os layer and 1 pt on each non-os layer (see Fig. 1) $n_b = 4N_{\text{os}} + (N_{\text{ps}} - N_{\text{os}})$
16	$[\hat{r}]_b := [\hat{M}]_j [\hat{r}]_b$ ($j = 1, \dots, n_f^2$)	apply $n_b \times n_b$ MVM
17	write n_b values $[\hat{r}]_b$ ($j = 1, \dots, n_f^2$) back to \hat{r}	write result back to original n_b locations
18	$[z]_k = \mathcal{F}_k^{-1}[\hat{r}]_k$ ($k = 1, \dots, N_{\text{ps}}$)	2D backward FFTs for each layer
19	$p = r$	initial search direction

$$\hat{r} := \sum_{j=1}^{n_f^2} P_j^T \hat{M}_j [r]_b, \quad [r]_b = P_j \hat{r}. \quad (7)$$

Each $n_b \times n_b$ block \hat{M}_j is efficiently computed offline as follows:

$$\hat{M}_j = \hat{Q}_j^{-1}, \quad \hat{Q}_j = P_j \hat{A}' P_j^T. \quad (8)$$

\hat{A}' can be expressed analytically under the usual FD modeling assumptions of infinite aperture and uniform measurement noise (spatial invariance) [24]. A derivation is provided in Appendix B. Figure 1 provides a graphic illustration of how Eq. (7) is implemented in practice. It involves a 2D FFT of each atmospheric screen, reading four points on each oversampled layer, and one point on each nonoversampled layer following the ordering illustrated in the figure (operator P_j), multiplying by a small precomputed Hermitian $n_b \times n_b$ matrix, and writing the result back to the original locations (operator P_j^T). This step is performed in parallel n_f^2 times for all $n_f \times n_f$ WFS Fourier plane grid points. Finally, it should

be noted that the final N_{ps} backward FFTs (line 18) return a real-valued phase vector.

Table 2 describes the FDPCG iterations, each iteration refining the approximate solution vector. For a total of i_{max} iterations, $(i_{\text{max}} + 1)$ A -mult operations and i_{max} applications of the preconditioner are required. The output solution is the turbulence estimate \hat{x}_{n-1} . An optional prediction step computing $\hat{x}_{n+1|n}$ can then be applied to \hat{x}_{n-1} (line 32). The two-step predictor takes the form of a spatial shift modeling frozen flow of magnitude and direction provided by an external estimate of the wind profile (not discussed in this paper).

Finally, Table 3 describes DM fitting and temporal filtering. Line 33 describes ray tracing to the aperture plane along N_{fit} fitting directions, line 34 computes the fitting right-hand-side vector, and line 35 yields finally the least-squares solution to the DM influence function cross-coupling fitting matrix system. The command vector is then typically temporally filtered via a finite difference equation (line 36), e.g., a first- or second-order low-pass filter.

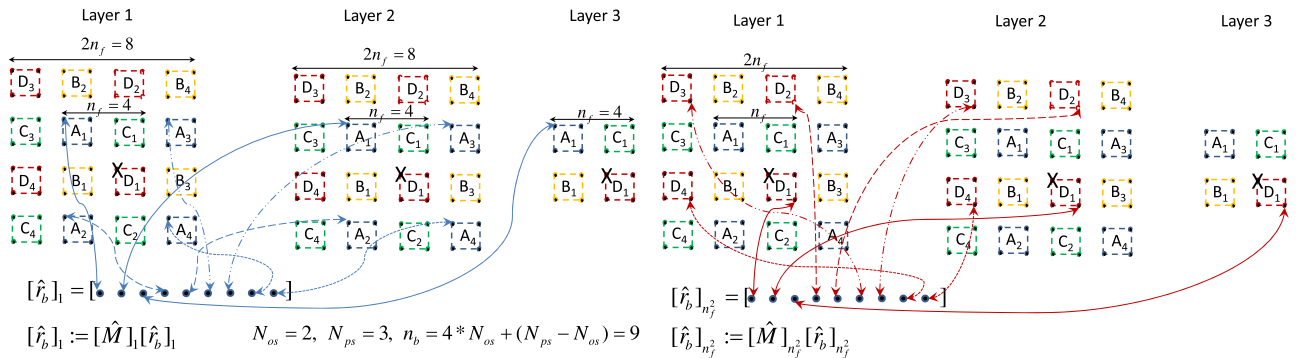


Fig. 1. Illustration of the key operations involved in the FD preconditioning step. The case $N_{\text{ps}} = 3, N_{\text{os}} = 2, n_f = 4$ (toy problem of order 2×2) is illustrated. The origin has been marked with an "X" symbol. For each point $j = 1, \dots, n_f^2$ of the $n_f \times n_f$ WFS Fourier plane, $n_b = 4N_{\text{os}} + (N_{\text{ps}} - N_{\text{os}})$ locations distributed across the N_{ps} atmospheric Fourier planes are read out, multiplied by a precomputed complex-valued $n_b \times n_b$ FD matrix \hat{M}_j , and written back to their original locations.

Table 2. FDPCG Tomography Algorithm Iterations and Output Approximate Solution

20	For $i = 1, \dots, i_{\max}$	loop of FDPCG iterations
21	$q = Ap$	imaged search direction (A -mult)
22	$\gamma = z^T r$	dot product
23	$\lambda = \gamma / (p^T q)$	dot product
24	$x := x + \lambda p$	solution update
25	If $i < i_{\max}$	if test on iteration number
26	$r := r - \lambda q$	residual update
27	$z = \text{Precon}(r, [\hat{M}]_1, \dots, [\hat{M}]_{n_f^2})$	preconditioning step
28	$\epsilon = z^T r / \gamma$	dot product
29	$p := \epsilon p + z$	search direction update
30	End	end of if test on iteration number; end of for loop on iteration number
31	$\hat{x}_{n-1} := x$	solution after i_{\max} iterations
32	$\hat{x}_{n+1 n} = H_v^2 \hat{x}_{n-1}$	optional two-step prediction using wind profile estimate (bilinear interpolation)

Table 5 in Appendix C summarizes the required 1D number of grid points for LGS MCAO observations with TMT at 60 deg zenith angle and for a 2 arcmin diameter fitting field of view. The atmospheric propagators and aperture-plane gradient operator are implemented on those square grids. This applies to both FDPCG and DKF. Regarding memory storage requirements, we use the value of 4 bytes per real and integer.

Table 6 in Appendix C provides a detailed computational cost analysis for three FDPCG iterations, which is the required number of iterations to reach optimal performance (lowest residual WFE). Reference has been made to the associated mathematical expressions of Table 1 (line numbers in brackets and bold). All propagators and the curvature operator are implemented on the grid, based on their regular stencils, and are therefore virtually memory-free operations. The only operator requiring sparse matrix storage is the DM interaction matrix. The WFS gradient matrices are stored as 18 weights per subaperture per WFS, and the FD preconditioner is stored as $n_f^2 n_b (n_b + 1) / 2$ complex-valued numbers. In summary, the estimated number of operations per frame ranges from 42 to 176 millions of multiplications and accumulations (MMACs) depending on the number of oversampled layers, and memory requirements range from 3 to 38 MB.

Lines 86 and 87 summarize the incremental WFE in quadrature with respect to the case $N_{\text{os}} = N_{\text{ps}}$ (FD3o6). Performance estimates were obtained from high-fidelity wave-optics simulations for observations at zenith and median Mauna Kea (0.55 arcsec at 500 nm) using LAOS, a MATLAB, high-fidelity, end-to-end AO simulation tool, particularly well suited for algorithm development and prototyping [25]. An overview of the code's LGS wave-optics capabilities can be found in [26]. We observe that there is 40 nm RMS variability in quadrature between FD3 and FD3o6, and that the two-step prediction with perfect wind profile knowledge reduces the WFE by ~ 16 nm, which is the estimated magnitude of the servo-lag error.

Table 3. DM Fitting and Temporal Filtering

33	$\psi = H \hat{x}_{n+1 n}$	ray tracing along N_{fit} fitting directions (bilinear interpolation)
34	$b_a = H_a^T W_0 \psi$	fitting right-hand-side vector
35	$\tilde{a}_{n+1} = (H_a^T W_0 H_a)^{\dagger} b_a$	unfiltered DM command
36	$a_{n+1} = (1 - g)a_n + g\tilde{a}_{n+1}$	filtered DM command (LPF)

B. DKF

Kalman filtering (KF), a key ingredient of linear quadratic Gaussian control, has been extensively investigated in the context of AO tomography by several authors (see, e.g., [27] and references therein). The Kalman filter operates recursively on the current measurement and previously computed turbulence estimate to produce a statistically optimal turbulence estimate for the next frame. It is derived from the linear “plant” model (1), supplemented with a turbulence temporal evolution model. We consider a first-order auto-regressive (AR1) turbulence temporal evolution model [28,29]:

$$x_n = H_v x_{n-1} + \nu_{n-1}, \quad (9)$$

where H_v is a bilinear interpolation matrix modeling a one-step frozen flow shift, and ν_n is an additive zero-mean turbulence boiling white noise term with spatial covariance matrix given by

$$C_\nu = C_x - H_v C_x H_v^T. \quad (10)$$

The algorithm is described in Table 4, and consists in computing the measurement signal produced by the previous frame tomography estimate $\hat{x}_{n-1|n-1}$, subtracting it from the pseudo-open-loop measurement (2), removing global TT, applying the precomputed Kalman gain, and updating $\hat{x}_{n-1|n-1}$. Application of the Kalman gain can be done either via SD convolutions (line 93), i.e.,

$$[[u]_k]_{p,q} = \sum_{i=1}^{N_{\text{igs}}} \sum_{p',q'=-n_K/2}^{n_K/2-1} ([[\tau_x]_{ki}]_{p-p',q-q'} [[\tilde{e}_x]_i]_{p',q'} + [[\tau_y]_{ki}]_{p-p',q-q'} [[\tilde{e}_y]_i]_{p',q'}), \quad (11)$$

or more efficiently via FFTs and component-wise multiplications by precomputed Fourier filters (line 94). The $2N_{\text{igs}}N_{\text{ps}}$ Fourier filters $[\hat{\tau}_{x(y)}]_{k,i}$ are obtained by solving offline the discrete algebraic Riccati equation [30] in the FD, assuming an infinite aperture and uniform measurement noise (spatial invariance). Mathematically, this offline computation can be expressed as follows:

$$\hat{C}_{t+1} = f(\hat{C}_t), \quad (12)$$

where t denotes the Riccati solver iteration number, $[\hat{C}_t]_{kl} = \mathcal{F}_k[C_t]_{kl}\mathcal{F}_l^{-1}$ is the Fourier representation of block k, l of

Table 4. DKF Tomography Algorithm Description

88	DKF Computations	Goal is to compute turb. est $\hat{x}_{n n}$ and subsequently a_{n+1} from $\hat{x}_{n-1 n-1}$ and from s_n^{OL}
89	$[\hat{s}_{n n-1}^{\text{OL}}]_i = \Gamma_i \sum_{k=1}^{N_{\text{ps}}} \hat{H}_{ik} [\hat{x}_{n-1 n-1}]_k$	measurement signal from tomography estimate $\hat{x}_{n-1 n-1}$
90	$e = s_n^{\text{OL}} - \hat{s}_{n n-1}^{\text{OL}}$	pseudo-open-loop error signal
91	$[\bar{e}_{x(y)}]_i = [e_{x(y)}]_i - [\omega_{x(y)}]_i^T [e]_i$	global tip/tilt removal with gradient weights $[\omega_{x(y)}]_i$
92	$u = K \bar{e}$	Kalman gain mult
93	$[u]_k = \sum_{i=1}^{N_{\text{igs}}} ([\tau_{x}]_{ki} \star [\bar{e}_x]_i + [\tau_{y}]_{ki} \star [\bar{e}_y]_i)$	SD convolutions on 2D arrays
94	$[\hat{e}_{x(y)}]_i = \mathcal{F}[\bar{e}_{x(y)}]_i$ $[\hat{u}]_k = \sum_{i=1}^{N_{\text{igs}}} ([\hat{\tau}_{x}]_{ki} \cdot \star [\hat{e}_x]_i + [\hat{\tau}_{y}]_{ki} \cdot \star [\hat{e}_y]_i)$ $[u]_k = \mathcal{F}^{-1}[\hat{u}]_k$	convolutions via 2D FFTs and 2D FD filter component-wise array mults
95	$\hat{x}_{n-1 n} = \hat{x}_{n-1 n-1} + u$	solution update
96	$\hat{x}_{n n} = H_v \hat{x}_{n-1 n}$	one-step prediction using wind profile estimate to obtain current frame turbulence estimate $\hat{x}_{n n}$ to be stored in memory
97	$\hat{x}_{n+1 n} = H_v \hat{x}_{n n}$	one-step prediction using wind profile est and ray tracing along N_{fit} directions
98	[B33–B36]	fitting right-hand-side vector, unfiltered command and final temporally filtered command

the covariance matrix C_t of the estimation error $x_t - \hat{x}_{t|t}$ at frame t , and the nonlinear function $f(\cdot)$ is given by

$$f(\hat{C}_t) = \hat{H}_v(I - \hat{K}_t \hat{G}') \hat{C}_t \hat{H}_v^{\star T} + \hat{C}_v, \quad (13)$$

where $\hat{G}' = \hat{\Gamma} \hat{H}$ is the Fourier representation of the infinite aperture, unobscured, approximation G' of the measurement matrix $G = \Gamma \tilde{H}$. An analytic expression of \hat{G}' is given in Appendix B. The FD Kalman gain at frame t , \hat{K}_t is obtained as

$$\hat{K}_t = \hat{C}_t \hat{G}'^{\star T} (\hat{G}' \hat{C}_t \hat{G}'^{\star T} + \sigma^2)^{-1}, \quad (14)$$

with σ^2 equal to the aperture-averaged measurement noise variance. We denote by \hat{K} the value of the Fourier representation of the Kalman gain matrix when the convergence stopping criteria has been reached. All blocks of \hat{K} are diagonal matrices:

$$[\hat{K}_{x(y)}]_{ki} = \text{diag}(\text{vec}([\hat{\tau}_{x(y)}]_{ki})), \quad (15)$$

where $[\hat{\tau}_{x(y)}]_{ki}$ are the 2D Fourier transforms of the 2D spatial convolution kernels $[\tau_{x(y)}]_{ki}$ for layer k and WFS i , and $\text{vec}(\cdot)$ raster scans a 2D array and stores entries in a long vector. Sample convolution kernels for DKF's SD implementation (line 93) are illustrated in Fig. 2.

Note that in order to guarantee convergence of the Riccati solver, the Fourier representation of the shift matrix, \hat{H}_v , has to be multiplied by a value slightly smaller than unity, which we chose to be equal to 0.99. Following the update step, the turbulence estimate $\hat{x}_{n-1|n}$ is propagated forward in time (line 96) via a one-step predictor, and the result, $\hat{x}_{n|n}$, is stored for next frame's update. A second one-step prediction is applied to yield the turbulence estimate for next frame given measurements up to the current frame, $\hat{x}_{n+1|n}$, which is then least-squares fitted to the DMs and optionally temporally filtered (lines 33–36 of Table 3) to balance noise propagation and servo-lag error.

Table 7 in Appendix C provides a detailed cost analysis for the case $N_{\text{os}} = 0$, N_{ps} with the Kalman gain multiplication implemented via FFTs. We observe that the cost is reduced compared to that of three FDPCG iterations by a factor ranging from 2.7 (no layers oversampled) to 3.8 (all layers oversampled). This arises from the fact that FDPCG requires $2 N_{\text{ps}}$ FFTs per iteration (i.e., $6 N_{\text{ps}}$ FFTs for three iterations), whereas DKF requires only $(2 N_{\text{igs}} + N_{\text{ps}})$ FFTs. Performance estimates are given at lines 118–120 and are also plotted in Fig. 3. All simulations used a first-order low pass filter (LPF) with smoothing parameter $g = 1/2$ as the temporal filter (line 36 in Table 3). We observe that despite its spatial invariance assumption, DKF provides a significantly reduced WFE compared to the performance delivered by

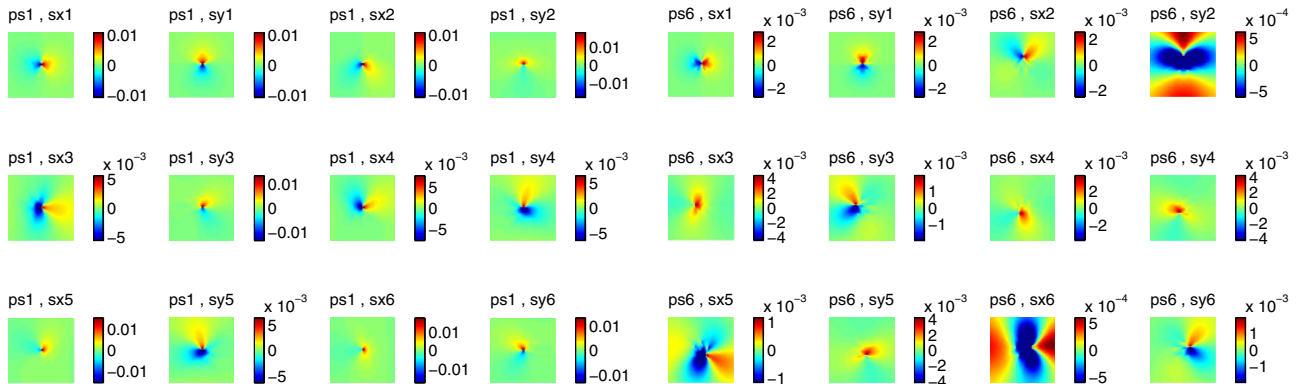


Fig. 2. Sample 60×60 convolution kernels for DKF's SD implementation for TMT ($N_{\text{igs}} = 6$ LGS WFSs) for the ground-level atmospheric layer ("ps1") (left panel) and top layer at 16 km altitude ("ps6") (right panel).

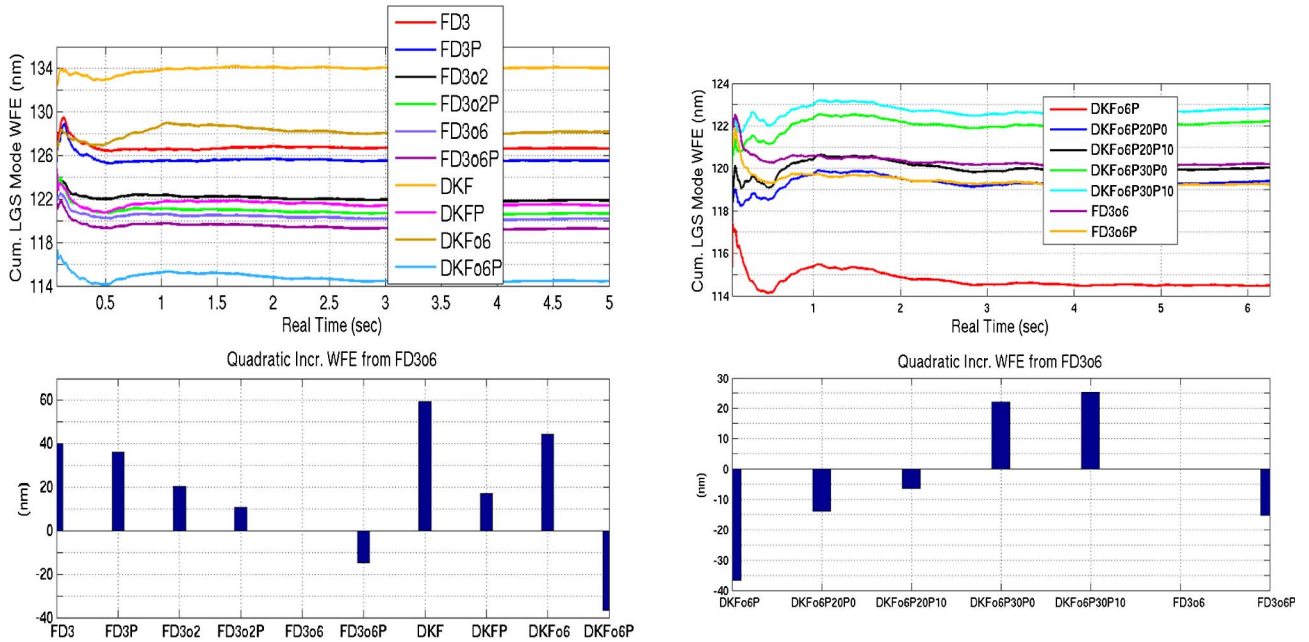


Fig. 3. FDPCG and DKF performance estimates (cumulative LGS mode WFE over a 17 arcsec square field of view, capturing all fundamental AO error budget terms) for zenith observations with TMT and median Mauna Kea seeing. Left is for algorithms without prediction (i.e., a zero wind speed assumption) and for algorithms with perfect wind profile knowledge (indicated with the letter “P”). Right is for algorithms with 20 or 30 deg wind direction error, and either 0% or 10% wind speed error on each atmospheric layer, denoted, respectively, as P20P0, P20P10, P30P0, and P30P10. Note that the y axes in the left and right figures scale differently.

three iterations of FDPCG, provided the wind profile is known to better than 20 deg wind direction and 10% wind speed error.

Finally, Table 8 in Appendix C reports operation counts when the Kalman gain is implemented via SD convolutions, which for TMT’s 60×60 LGS MCAO system, is significantly more costly than the FD implementation via FFTs.

3. CONCLUSIONS

We have discussed the performance and computational cost of an innovative wavefront reconstruction algorithm for wide-field LGS AO. The algorithm was proposed in [17] and consists of a distributed (spatially invariant) Kalman filter (DKF), which can be implemented in real time by means of massively parallelizable, low-memory, localized SD convolution kernels, or more efficiently via FFTs applied to the high-order WFS measurements. We have evaluated the performance of this algorithm in end-to-end, high-fidelity, Monte Carlo simulations of NFIRAOS, the MCAO system under design for TMT, and have compared it against the previously introduced FDPCG algorithm, whose cost scales as $N \log(N)$, where N denotes the dimensionality of the tomography space. If the Kalman gain multiplication is implemented via FFTs, DKF’s cost also scales as $N \log(N)$ and is reduced compared to the cost of three FDPCG iterations by a factor ranging from 2.7 (no layers oversampled) to 3.8 (all layers oversampled), whereas the SD convolution implementation is significantly more expensive. Despite its spatial invariance assumption, DKF ranks as a better algorithm in terms of turbulence correction, provided the wind profile is known to better than about 10% wind speed accuracy and 20 deg wind direction accuracy. Due to its overall simplicity, noniterative nature, and massively parallel implementation requiring minimal interprocessor communication, DKF appears to be particularly well suited for real-time implementation on inexpensive off-the-shelf GPUs.

Several approximations enter into the elaboration of DKF. First, the algorithm does not account for variation of WFS measurement noise arising from nonuniform spot elongation in the pupil. This is a limitation of the method that requires spatially invariant operators. However, the resulting filter is still capable of outperforming FDPCG provided an accurate wind profile estimate is available and incorporated in the algorithm. Second, the infinite pupil hypothesis leads to a Kalman gain that does not account for edge effects. First attempts to mitigate this assumption have shown that performance could be increased with appropriate modifications of the filters. How to optimize these modifications while maintaining closed-loop stability remains an open question that will be revisited in the future.

In this paper, DKF was implemented with an AR1 turbulence model. This simple-minded model is appropriate in our context of very large dimensions. It is jointly used with a first-order LPF on the command vector, which leads to an implicit second-order turbulence model [28], thereby improving performance (higher-order models, e.g., AR2, are known to improve performance [27], suggesting that turbulence is probably closer to an AR2 than an AR1). Regarding performance, it should be pointed out that the choice of turbulence model and filtering applied to the command vector determine the ultimate performance level. In this regard, other combinations than AR1 + LPF could be envisioned for DKF without significant extra computational cost.

Upcoming developments will be targeted toward the analysis of a full FD implementation with reduced cost, wind profiling via slope detection and ranging (SLODAR), a technique already implemented on the Gemini South LGS MCAO system [31], and the analysis of alternative turbulence models while keeping the computational and performance attractiveness of the algorithm.

APPENDIX A: LAPLACIAN MATRIX SCALING

Each block k of the Laplacian regularization matrix is written as $L_k = c_k \bar{L}_k$, where \bar{L}_k as the following normalized stencil:

$$\text{stencil}(\bar{L}_k) = \begin{bmatrix} 0 & \omega_3 & 0 \\ \omega_4 & \omega_1 & \omega_2 \\ 0 & \omega_5 & 0 \end{bmatrix}, \quad (\text{A1})$$

where the central weight is equal to -4 , and the surrounding weights are equal to 1. The stencil width is $2\Delta_k$, and the Cartesian coordinates of the stencil points will be denoted $p_{k,j}$. The scaling constant c_k is chosen such that $\langle x_k^T C_{x_k}^{-1} x_k \rangle = c_k^2 \langle x_k^T \bar{L}_k^2 x_k \rangle$, where x_k denotes the zero-mean turbulence phase for layer k , C_{x_k} denotes the spatial covariance matrix of x_k , and $\langle \cdot \rangle$ denotes ensemble average. Denoting by N_k the number of phase points on layer k , we have

$$c_k^2 = N_k / \langle x_k^T \bar{L}_k^2 x_k \rangle. \quad (\text{A2})$$

Since turbulence is stationary and isotropic, we have

$$\langle x_k^T \bar{L}_k^2 x_k \rangle = N_k \left\langle \left(\sum_{j=1}^5 \omega_j [x_k]_j \right)^2 \right\rangle. \quad (\text{A3})$$

Since the five weights ω_j sum to zero, Eq. (A3) can be rewritten as a function of phase differences, which can then be expressed in terms of the Kolmogorov phase structure function [32]:

$$\begin{aligned} \left\langle \left(\sum_{j=1}^5 \omega_j [x_k]_j \right)^2 \right\rangle &= -\frac{1}{2} \sum_{j,j'=1}^5 \omega_j \omega_{j'} \langle ([x_k]_j - [x_k]_{j'})^2 \rangle \\ &= 3.44 r_{0,k}^{-5/3} \beta_k, \end{aligned} \quad (\text{A4})$$

where

$$\beta_k = - \sum_{j,j'=1}^5 \omega_j \omega_{j'} \|p_{k,j} - p_{k,j'}\|^{5/3} = (2 - 2^{-1/6} - 2^{-1/3}) \Delta_k^{5/3}. \quad (\text{A5})$$

Combining Eqs. (A2)–(A5) yields

$$c_k = 0.24 (r_{0,k} / \Delta_k)^{5/6}. \quad (\text{A6})$$

APPENDIX B: FOURIER REPRESENTATION OF TOMOGRAPHY MATRIX

The symmetric positive definite approximation A' of A rests on the following fundamental approximations:

1. A' is constructed on oversized $N_{\text{fft}}/\alpha_k \times N_{\text{fft}}/\alpha_k$ computational grids extending up to twice the size of the metapupils at the different atmospheric layers. See Table 6 in Appendix C for sample values for TMT.

2. A' approximates spatially varying LGS WFS measurement noise arising from LGS spot elongation as aperture-plane averaged noise.

3. A' incorporates an oversized $N_{\text{fft}} \times N_{\text{fft}}$ aperture-plane grid ($N_{\text{fft}} \geq 2D/\Delta_0$) on which $N_{\text{fft}}^2 \times N_{\text{fft}}^2$ gradient matrices $\Gamma'_{x(y)}$ are constructed as follows:

$$\Gamma'_{x(y)} = \Omega \bar{\Gamma}'_{x(y)}, \quad (\text{B1})$$

where the diagonal mask matrix Ω is expressed as

$$\begin{aligned} \Omega &= \text{diag}(\text{vec}(\omega)), \\ [\omega]_{jj'} &= \begin{cases} m_0 = n_{\text{sa}}/(N_{\text{fft}}/2)^2 & \text{if } j, j' \text{ are even, else} \\ 0 & \end{cases}, \end{aligned} \quad (\text{B2})$$

and the gradient matrices $\bar{\Gamma}'_{x(y)}$ have uniform stencils throughout the grid given by

$$\begin{aligned} \text{stencil}(\bar{\Gamma}'_x) &= \frac{1}{(2\Delta_0)} \begin{bmatrix} -1/4 & 0 & 1/4 \\ -1/2 & 0 & 1/2 \\ -1/4 & 0 & 1/4 \end{bmatrix}, \\ \text{stencil}(\bar{\Gamma}'_y) &= -\text{stencil}(\bar{\Gamma}'_x)^T. \end{aligned} \quad (\text{B3})$$

The $\text{vec}(\cdot)$ operation in Eq. (B2) raster scans a 2D array and stores its entries in a long vector. The role of the mask matrix Ω is twofold: it takes into account the fact that the aperture-plane grid is oversampled and therefore only half of the grid points in each direction are physical measurement grid points, and it de-weights the physical measurements by an amount equal to m_0 in order to preserve the norm of the measurement vector for an input TT mode.

We thus have

$$[A']_{kl} = \sum_{i=1}^{N_{\text{igs}}} [G']_{ik}^T \sigma_i^{-2} [G']_{il} + \delta_{k,l} L_k^2, \quad (\text{B4})$$

where σ_i^{-2} denotes the inverse of LGS WFS i aperture-plane averaged measurement noise variance, and

$$[G']_{ik} = \begin{bmatrix} \Gamma'_x \\ \Gamma'_y \end{bmatrix} [\tilde{H}]_{ik}. \quad (\text{B5})$$

Periodic boundary conditions and the cone coordinate system are enforced on all matrices in order to yield shift-invariant matrices with diagonal FD representations.

The FD representation of Eq. (25) is expressed as follows:

$$[\hat{A}']_{kl} = \mathcal{F}_k [A']_{kl} \mathcal{F}_l^{-1} = \sum_{i=1}^{N_{\text{igs}}} [\hat{G}']_{ik}^{\star T} \sigma_i^{-2} [\hat{G}']_{il} + \delta_{k,l} \hat{L}_k^2, \quad (\text{B6})$$

where \star denotes the complex conjugate, and

$$\begin{aligned} [\hat{G}']_{ik} &= \begin{bmatrix} \hat{\Gamma}'_x \\ \hat{\Gamma}'_y \end{bmatrix} [\hat{\tilde{H}}]_{ik}, & \hat{\tilde{H}}_{ik} &= \mathcal{F}_0 [\tilde{H}]_{ik} \mathcal{F}_k^{-1}, \\ \hat{\Gamma}'_{x(y)} &= \mathcal{F}_0 \Gamma'_{x(y)} \mathcal{F}_0^{-1}, \end{aligned} \quad (\text{B7})$$

where \mathcal{F}_0 denotes the aperture-plane Fourier matrix.

1. Fourier Representation of Propagation Matrix

For oversampled layers ($\alpha_k = 1$), the propagation operator is simply a shift (on account of the cone coordinate and the oversampling of the aperture plane), which has diagonal representation in Fourier space. For nonoversampled layers, the operator corresponds to a combined shift and upsampling, and is therefore nondiagonal (four nonzero elements per column). We thus have

$$[\hat{H}]_{ik} = \begin{cases} \text{diag}(\text{vec}(\Lambda_{ik})) & \text{if } \alpha_k = 1, \text{ else} \\ \text{diag}(\text{vec}(\Theta_0))T \text{diag}(\text{vec}(\Lambda_{ik})), & \end{cases} \quad (\text{B8})$$

where the $(N_{\text{fft}}/\alpha_k) \times (N_{\text{fft}}/\alpha_k)$ array Λ_{ik} is given by the following expression:

$$\Lambda_{ik} = \exp(-i\pi h_k [f_{k,x}\theta_i^x + f_{k,y}\theta_i^y]), \quad (\text{B9})$$

where $\bar{i} = \sqrt{-1}$, and $f_{k,x(y)}$ are the spatial frequency arrays for layer k (columns of $f_{k,x}$ are identical and $f_{k,y} = f_{k,x}^T$). Matrix T denotes the $(0,1)$ valued $N_{\text{fft}}^2 \times N_{\text{fft}}^2/4$ periodic extension matrix (four nonzero elements per column), which periodically extends each quadrant A, B, C, D of the WFS Fourier plane, i.e.,

$$\begin{bmatrix} D & B & D & B \\ C & A & C & A \\ D & B & D & B \\ C & A & C & A \end{bmatrix} = \text{array}(Tv), \quad v = \text{vec}\left(\begin{bmatrix} A & C \\ B & D \end{bmatrix}\right), \quad (\text{B10})$$

where the $\text{array}(\cdot)$ operation performs the inverse of the $\text{vec}(\cdot)$ operation. Finally, the aperture-plane upsampling array Θ_0 is derived from the contributions of coincident, mid x -vertex, mid y -vertex, and central points of 3×3 unit cells, resulting in the following expression:

$$\Theta_0 = \frac{1}{2}(1 + \cos(\varphi_{0,x}) + \cos(\varphi_{0,y}) + \cos(\varphi_{0,x})\cos(\varphi_{0,y})), \quad (\text{B11})$$

where $\varphi_{0,x(y)} = 2\pi\Delta_0 f_{0,x(y)}$ and $f_{0,x(y)}$ are the $N_{\text{fft}} \times N_{\text{fft}}$ aperture-plane spatial frequency arrays.

2. Fourier Representation of Laplacian Matrix

The Laplacian matrix with periodic boundary conditions has a block circulant with circulant block structure [33] and is therefore diagonal in Fourier space:

$$\hat{L}_k = \text{diag}(\text{vec}(\hat{\psi}_k)), \quad \hat{\psi}_k = 4c_k(\sin^2(\varphi_{k,x}/2) + \sin^2(\varphi_{k,y}/2)), \quad (\text{B12})$$

where $\varphi_{k,x(y)} = 2\pi\Delta_k f_{k,x(y)}$ and the scaling constant c_k is given in Eq. (A6).

3. Fourier Representation of Aperture-Plane Gradient Matrix

Since the aperture plane is oversampled, the FD representation of the gradient matrix is the product of a diagonal matrix and a sparse convolution matrix:

$$\hat{\Gamma}_{x(y)}' = \hat{\Omega} \text{diag}(\text{vec}(\hat{g}_{x(y)})), \quad (\text{B13})$$

where $\hat{\Omega} = BCCB(\hat{\omega})$ is a convolution matrix whose $N_{\text{fft}} \times N_{\text{fft}}$ generator $\hat{\omega}$ (convolution kernel) is the 2D Fourier transform of the aperture mask array given in Eq. (B2). Such a kernel consists of four discrete delta functions located at the vertices of the $(-, -)$ quadrant of the aperture-plane Fourier plane. These delta functions are positive along the diagonal and negative along the antidiagonal, and have all the same magnitude. The FD gradient arrays $\hat{g}_{x(y)}$ are found from Eq. (24), together with the following property:

$$\phi_{jj'} = \text{vec}(e^{i(j\varphi_{0,x} + j'\varphi_{0,y})})^T \text{vec}(\hat{\phi}), \quad (\text{B14})$$

where $\hat{\phi}$ denotes the 2D Fourier transform of array ϕ , and j, j' grid point indices, resulting in the following expression:

$$\begin{aligned} \hat{g}_x &= \frac{e^{-i(\varphi_{0,x} + \varphi_{0,y})}}{8\Delta_0} [e^{2i\varphi_{0,y}} + 2e^{i(2\varphi_{0,y} + \varphi_{0,x})} \\ &\quad + e^{2i(\varphi_{0,x} + \varphi_{0,y})} - 1 - 2e^{i\varphi_{0,x}} - e^{2i\varphi_{0,x}}], \\ \hat{g}_y &= -\hat{g}_{x \rightarrow y}. \end{aligned} \quad (\text{B15})$$

APPENDIX C: FDPG AND DKF COMPUTATIONAL BUDGETS

Table 5. Required Atmospheric and Aperture-Plane Grid Sizes for LGS MCAO Observations with TMT at 60 deg Zenith Angle and for a 2 arcmin Diameter Fitting Field of View^a

37	B	C	D	E
38	1D nb of phase points per layer at 60 deg zenith angle for 2 arcmin diameter fitting field (square boxes enclosing metapupils)			
		FD3 or DKFP	FD3o2	FD3o6 or DKFo6P
39	$n1$ (layer 1)	61	123	123
40	$n2$ (layer 2)	71	143	143
41	$n3$ (layer 3)	80	80	161
42	$n4$ (layer 4)	90	90	179
43	$n5$ (layer 5)	110	110	221
44	$n6$ (layer 6)	122	122	245
45	$n0$ (aperture-plane grid for Γ_i)		123	
46	Mem 1e6 real numbers		3.81 MB = $1e6/(1024)^2$	

^aSuch grid sizes are large enough to accommodate LGSs as low as 80 km in altitude.

Table 6. Computation Requirements for Three FDPCG Iterations for the Case $N_{os} = 0, 2, N_{ps}$, Denoted, Respectively, as FD3, FD3o2, and FD3o6^a

47	FDPCG Computations				
48		FD3	FD3o2	FD3o6	FD3o6 Formulas
49	Nlgs		6		Nb. of LGSs
50	Nps		6		Nb. of estimated phase screens
51	Nos	0	2	6	Nb. of layers oversampled
52	Ndm		2		Nb. of DMs
53	Ng/1e6		0.03		Nb. of gradients
54	nnz(W)/Ng		2.00		Noise weighting
55	Cost W mult		0.06		$C54 * C53$
56	nnz(Ga)/(Ng * Ndm)		20.18		DM interaction matrix (bicubic IF)
57	Cost Ga mult		1.25		$C56 * C53 * C52$
58	Mem Ga (Compressed Sparse Row)		9.66		$(2 * C57 + C53) * C46$
59	Niter		3		Nb. of CG iterations
60	N/1e6	0.05	0.08	0.20	$(E39^2 + \dots + E44^2)/1e6$
61	Nfft	128	mixed	256	1D nb. of phase pts on FFT grids
62	Cost of single Nfft x Nfft FFT	0.23	NA	1.05	
63	nnz(Gamma)/Ng		8.07		aperture-plane WFS grad matrix
64	Mem Gamma		0.95		$C63 * C53 * C46$
65	N0/1e6		0.02		$C45^2/1e6$
66	nnz(Htilde)/(N0 * Nlgs * Nps)	3.39	3.12	2.93	ray tracing LGS-to-aperture
67	Cost(Gamma * Htilde)mult	2.10	1.95	1.84	$C63 * C53 + E66 * C65 * C49 * C50$
68	Cost open-loop grads [B3]		1.28		$C57 + C53$
69	Cost tip/tilt bias removal [B5]		0.06		$2 * C53$
70	Cost tomo rhs vec [B6]	2.16	2.01	1.91	$E67 + C55$
71	nnz(L)/N	4.96	4.97	4.97	Laplacian matrix (regularization)
72	Cost L mult	0.25	0.38	0.38	$E71 * D60$
73	Cost tomo matrix mult/iter [B8]	4.75	4.73	4.52	$2 * E67 + C55 + 2 * E72$
74	Cost precondition FFTs/iter [B14, B18]	2.75	6.03	12.58	$2 * C50 * E62$
75	Cost precondition MVM/iter [B16]	2.36	9.44	37.75	$C61^2 * 4 * (4 * E51)^2/1e6$
76	Cost precondition/iter [B13]	5.11	15.47	50.33	$E74 + E75$
77	Cost tomo matrix mult [B8,B21]	19.00	18.92	18.06	$(C59 + 1) * E73$
78	Cost precondition [B13,B27]	15.34	46.40	150.99	$C59 * E76$
79	Mem precondition	2.63	9.75	37.50	$E75/4 * C46 * (1 + 1/(4 * E51))$
80	Cost dot prods, vec adds [B8,B22–B24,B26,B28,B29]	0.80	1.23	1.21	$E60 * 6 * (C59 - 2)$
81	Cost total tomo	38.64	69.91	173.52	$C68 + C69 + E70 + E77 + E78 + E80$
82	Nfit		9		Nb. of fitting directions
83	nnz(H)/(N0 * Nfit * Nps)	3.70	3.49	3.49	ray tracing fitting field-to-aper
84	Cost H mult [B33]	3.02	2.85	2.85	$E83 * C65 * C82 * C50$
85	Prefitting Cost grand total	41.66	72.76	176.38	$E81 + E84$
86	Quadratic Incr. from FD3o6 (nm) with no wind profile knowledge ($v = 0$)	40.00	20.00	0.00	
87	Quadratic Incr. WFE from FD3o6 (nm) with perfect wind profile knowledge	35.00	10.00	–16.00	Negative sign indicates reduced WFE

^aThe number of operations is given in units of MMACs per frame, and memory requirements in units of megabytes (MB). The last column contains the formulas used for FD3o6 (letters refer to columns and numbers to rows).

Table 7. DKF Tomography Algorithm Computational Cost Analysis for the Case of the Kalman Gain Multiplication Implemented via FFTs for the Case $N_{os} = 0$, N_{ps} , Denoted, Respectively, as DKFP and DKFo6P^a

99	DKF Comps with Kalman Gain Mult Implemented via FFTs			
100		DKFP	DKFo6P	DKFo6P Formulas
101	N/1e6	0.05	0.20	E60
102	Nfft	128	256	
103	N0/1e6		0.02	C65
104	Cost open-loop grads [B3]		1.28	C68 closed-loop grads + DM grads
105	Cost tomo grads [B89]	2.10	1.84	E67 Gamma * Htilde * tomo estimate
106	Cost grad est. error [B90]	3.41	3.16	D105 + C104 + C53
107	Cost tip/tilt bias removal [B91]		0.06	C69
108	Nb. of convolution kernels		72	C50 * C49 * 2
109	Cost of Kalman gain mult FFTs [B94]	4.13	18.87	E62 * (2 * C49 + C50) forward + backward FFTs
110	Cost of Kalman gain mult FD filter mults [B94]	4.72	18.87	C108 * D102 ² * 4/1e6 (complex-valued mults)
111	Total Cost Kalman gain mult	8.85	37.75	D110 + D109
112	Mem FD Kalman gain	9.00	36.00	2 * D102 ² * C108 * C46/1e6 (complex-valued)
113	Cost phase vector add [B95]	0.10	0.40	2 * D101 tomo estimate
114	Cost shift matrix mult [B96,B97]	0.40	1.62	2 * 4 * D101 predictions
115	Cost total tomo	14.00	43.87	D114 + D111 + C107 + D106 + C104
116	Cost H mult [B33]	3.02	2.85	E84
117	Prefitting Cost grand total	17.02	46.72	D115 + D116
118	Quadratic Incr. WFE from FD3o6 (nm) with perfect wind profile knowledge	18.00	-38.00	Negative sign indicated reduced WFE
119	Quadratic Incr. WFE from FD3o6 (nm) with no wind profile knowledge ($v = 0$)	60.00	42.00	
120	Quadratic Incr. WFE from FD3o6 (nm) with 20 deg wind direction error and 10% wind speed error	41.00	-9.00	

^aThe last column contains the formulas used for DKFo6P (letters refer to columns and numbers to rows).

Table 8. DKF Tomography Algorithm Computational Cost Analysis for the Case of the Kalman Gain Multiplication Implemented via SD Convolutions, for the Case $N_{os} = 0$, N_{ps} , Denoted, Respectively, as DKFP-SD and DKFo6P-SD^a

121	DKF Comps with Kalman Gain Mult Implemented via SD Convolutions			
122		DKFP-SD	DKFo6P-SD	DKFo6P-SD Formulas
123	nK	60	120	1D nb of points in convolution kernels
124	Ng/1e6	0.03	0.12	4 * C124
125	Cost grad est. error [B90]	3.41	3.16	D106
126	Cost tip/tilt bias removal [B91]		0.06	C69
127	Cost SD Kalman gain mult [B93]	669.60	10713.60	D123 ² * C50 * D124
128	Cost SD Kalman gain mult/Cost FD Kalman gain mult	75.68	283.81	D127/D111
129	Mem SD Kalman gain	0.99	3.96	D123 ² * C108 * C46/1e6

^aThe last column contains the formulas used for DKFo6P-SD (letters refer to columns and numbers to rows).

ACKNOWLEDGMENTS

The work of the second author has been supported by the Netherlands Organization for Scientific Research (NWO) and the Marie Curie COFUND Action by means of a Rubicon grant. This work has also been supported by University Paris 13 and by the French National Research Agency (ANR) through the project CHAPERSON ANR-09-BLAN-0162-01. The authors gratefully acknowledge the support of the TMT partner institutions. They are the Association of Canadian Universities for Research in Astronomy (ACURA), the California Institute of Technology, the University of California, the National Astronomical Observatory of Japan, the National Astronomical Observatories of China and their consortium partners, and the Department of Science and Technology of India and their supported institutes. This work was supported as well by the Gordon and Betty Moore Foundation, the

Canada Foundation for Innovation, the Ontario Ministry of Research and Innovation, the National Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada, the British Columbia Knowledge Development Fund, the Association of Universities for Research in Astronomy (AURA), and the U.S. National Science Foundation.

REFERENCES

1. R. Conan, B. Espeland, M. A. Van Dam, and A. H. Bouchez, "Design of the laser tomography adaptive optics system for the giant Magellan telescope," presented at the Second International Conference on Adaptive Optics for Extremely Large Telescopes, British Columbia, Canada, 25–30 September 2011.
2. T. Fusco, S. Meimon, N. Thatte, H. Schnettler, Y. Clenet, M. Cohen, J. Paufique, P. Ammans, F. Clarke, J. L. Dournaux, M. Ferrari, D. Gratadour, N. Hubin, P. Jagourel, V. Michau, C. Petit,

- and M. Tecza, "Laser tomographic AO system for an integral field spectrograph on the E-ELT: the ATLAS project," presented at the Second International Conference on Adaptive Optics for Extremely Large Telescopes, British Columbia, Canada, 25–30 September 2011.
3. G. Herriot, D. Andersen, J. Atwood, P. Byrnes, C. Boyer, K. Caputo, C. Correia, J. Dunn, B. Ellerbroek, J. Fitzsimmons, L. Gilles, P. Hickson, A. Hill, J. Pazder, V. Reshetov, M. Smith, J. P. Veran, L. Wang, and I. Wevers, "NFIRAOS: multi conjugate AO system for TMT," presented at the Second International Conference on Adaptive Optics for Extremely Large Telescopes, British Columbia, Canada, 25–30 September 2011.
 4. E. Diolaiti, I. Foppiani, J. M. Conan, C. R. Butler, R. I. Davies, A. Baruffolo, M. Bellazzini, G. Bregoli, P. Ciliegi, G. Cosentino, B. Delabre, T. Fusco, N. Hubin, M. Lombini, E. Marchetti, C. Petit, C. Robert, and L. Schreiber, "The E-ELT multi conjugate adaptive optics module," presented at the Second International Conference on Adaptive Optics for Extremely Large Telescopes, British Columbia, Canada, 25–30 September 2011.
 5. B. Ellerbroek and F. Rigaut, "Methods for correcting tilt anisoplanatism in laser guide star based multiconjugate adaptive optics," *J. Opt. Soc. Am. A* **18**, 2539–2547 (2001).
 6. T. Pfrommer and P. Hickson, "Properties and dynamics of mesospheric sodium and the impact on sodium LGS AO systems," presented at the Second International Conference on Adaptive Optics for Extremely Large Telescopes, British Columbia, Canada, 25–30 September 2011.
 7. L. Gilles and B. Ellerbroek, "Split atmospheric tomography using laser and natural guide stars," *J. Opt. Soc. Am. A* **25**, 2427–2435 (2008).
 8. L. Gilles, L. Wang, and B. Ellerbroek, "Minimum variance split tomography for laser guide star adaptive optics," *Eur. J. Control* **17**, 327–334 (2011).
 9. L. Wang, L. Gilles, and B. Ellerbroek, "Analysis of the improvement in sky coverage for multiconjugate adaptive optics systems obtained using minimum variance split tomography," *Appl. Opt.* **50**, 3000–3010 (2011).
 10. T. Fusco, J. M. Conan, G. Rousset, L. M. Mugnier, and V. Michau, "Optimal wave-front reconstruction strategies for multi-conjugate adaptive optics," *J. Opt. Soc. Am. A* **18**, 2527–2538 (2001).
 11. Q. Yang, C. Vogel, and B. Ellerbroek, "Fourier domain preconditioned conjugate gradient algorithm for atmospheric tomography," *Appl. Opt.* **45**, 5281–5293 (2006).
 12. C. Vogel and Q. Yang, "Fast optimal wavefront reconstruction for multi-conjugate adaptive optics using the Fourier domain preconditioned conjugate gradient algorithm," *Opt. Express* **14**, 7487–7498 (2006).
 13. E. Thiébaud and M. Tallon, "Fast minimum variance wavefront reconstruction for extremely large telescopes," *J. Opt. Soc. Am. A* **27**, 1046–1059 (2010).
 14. M. Tallon, C. Béchet, I. Tallon-Bosc, M. Le Louarn, E. Thiébaud, R. Clare, and E. Marchetti, "Performance of MCAO on the E-ELT using the fractal iterative method for fast atmospheric tomography," presented at the Second International Conference on Adaptive Optics for Extremely Large Telescopes, British Columbia, Canada, 25–30 September 2011.
 15. R. Ramlau and M. Rosenstainer, "An efficient solution to the atmospheric turbulence tomography problem using Kaczmarz iteration," *Inverse Probl.* **28**, 095004 (2012).
 16. R. Fraanje, J. Rice, M. Verhaegen, and N. Doelman, "Fast reconstruction and prediction of frozen flow turbulence on structured Kalman filtering," *J. Opt. Soc. Am. A* **27**, A235–A245 (2010).
 17. P. Massioni, C. Kulcsár, H. F. Raynaud, and J. M. Conan, "Fast computation of an optimal controller for large-scale adaptive optics," *J. Opt. Soc. Am. A* **28**, 2298–2309 (2011).
 18. L. Wang and B. Ellerbroek, "Computer simulations and real-time control of ELT AO systems using graphical processing units," *Proc. SPIE* **8447**, 844723 (2012).
 19. "AO tomography workshop 2012: presentations and discussions," available online <http://www.dur.ac.uk/cfai/adaptiveoptics/aotw12/agenda/>.
 20. C. R. Vogel, "Sparse matrix methods for wavefront reconstruction revisited," *Proc. SPIE* **5490**, 1327–1335 (2004).
 21. P. Piatrou and M. Roggemann, "Performance study of Kalman filter controller for multiconjugate adaptive optics," *Appl. Opt.* **46**, 1446–1455 (2007).
 22. L. Gilles and B. Ellerbroek, "Constrained matched filtering for extended dynamic range and improved noise rejection for Shack–Hartmann wavefront sensing," *Opt. Lett.* **33**, 1159–1161 (2008).
 23. B. Ellerbroek, "Efficient computation of minimum-variance wave-front reconstructors using sparse matrix techniques," *J. Opt. Soc. Am. A* **19**, 1803–1816 (2002).
 24. A. Tokovinin and E. Viard, "Limiting precision of tomographic phase estimation," *J. Opt. Soc. Am. A* **18**, 873–882 (2001).
 25. L. Gilles and B. Ellerbroek, "LAOS: linear adaptive optics simulator," Thirty Meter Telescope Observatory Corporation, Document-7988, TMT.AOS.TEC.07.003 (2007).
 26. L. Gilles, B. Ellerbroek, and J. P. Veran, "Laser guide star multi-conjugate adaptive optics performance of the thirty meter telescope with elongated beacons and matched filtering," *Proc. SPIE* **6272**, 627236 (2006).
 27. C. Kulcsár, H.-F. Raynaud, C. Petit, and J.-M. Conan, "Minimum variance prediction and control for adaptive optics," *Automatica* **48**, 1939–1954 (2012).
 28. C. Kulcsár, H. F. Raynaud, C. Petit, J. M. Conan, J. M. Conan, and P. V. de Lesegno, "Optimal control, observers and integrators in adaptive optics," *Opt. Express* **14**, 7464–7564 (2006).
 29. B. Le Roux, J. M. Conan, C. Kulcsár, H. F. Raynaud, L. Mugnier, and T. Fusco, "Optimal control law for classical and multiconjugate adaptive optics," *J. Opt. Soc. Am. A* **21**, 1261–1276 (2004).
 30. B. D. O. Anderson and J. B. Moore, *Optimal Control, Linear Quadratic Methods* (Prentice-Hall, 1990).
 31. A. Cortes, B. Neichel, A. Guesalaga, J. Osborn, F. Rigaut, and D. Guzman, "Atmospheric turbulence profiling using multiple laser star wavefront sensors," <http://arxiv.org/abs/1210.1999> (2012).
 32. M. Roggemann and B. Welsh, *Imaging Through Turbulence* (CRC, 1996).
 33. C. Vogel, *Computational Methods for Inverse Problems* (SIAM, 2002).